

Uma das formas que os elementos de rede (tipicamente routers) têm para controlar as sobrecargas de pacotes nos buffers é aplicando algoritmos de gestão de filas de espera.

- FIRST IN FIRST OUT (FIFO)
- Priority Queuing – PQ
- Class Based Queuing – CBQ Round Robin
- Weighted Fair Queuing –(WFQ)

Necessidade do Escalonamento:

- A partilha de recursos em redes de comunicação e, em particular, a adopção de estratégias de multiplexagem estatística, originam situações de contenção, devido à competição pela utilização de recursos
- Em redes que suportam integração de serviços torna-se necessário providenciar QoS diferenciada por categorias de serviço (ou classes de tráfego), oferecendo garantias de desempenho a aplicações críticas e ao mesmo tempo permitindo uma partilha de recursos de acordo com critérios de equidade
- Os algoritmos de escalonamento (scheduling algorithms) são um componente essencial para atingir estes objectivos, uma vez que determinam as disciplinas de serviço a aplicar a fluxos aceites pela rede
- Um algoritmo de escalonamento realiza duas funções:
 - decide a ordem de serviço de fluxos em competição (pedidos de serviço)
 - gere as filas de pedidos de serviço

Objetivos

- Os algoritmos de escalonamento são usados em qualquer sistema ou camada protocolar em que ocorra contenção por recursos, mas em especial em elementos de rede (routers, comutadores, etc.)
- Desempenham um papel importante na provisão de diferentes níveis de QoS a diferentes aplicações, permitindo controlo diferenciado de atraso, largura de banda ou taxa de perdas:
 - aplicações com requisitos de tempo real necessitam de garantias absolutas
 - aplicações best-effort não necessitam de quaisquer garantias, mas é desejável que a partilha de recursos entre estas aplicações seja feita de forma equitativa (fair)
- Em geral pretende-se que os algoritmos de escalonamento sejam simples (fáceis de implementar), tratem de forma equitativa tráfego best-effort e garantam ao restante tráfego os níveis de desempenho negociados e facilitem os mecanismos associados de controlo de admissão de ligações

Priority Queuing – PQ:

O algoritmo de fila de espera por prioridade (Priority Queuing - PQ) foi projetado para dar maior prioridade a determinado tráfego de dados que exigem certa urgência de processamento (sensíveis aos atrasos na fila de espera).

Class Based Queuing - CBQ

O algoritmo de fila de espera baseado em classes (Class Based Queuing - CBQ) foi projetado para permitir que várias aplicações, com especificações de largura de banda mínima ou sensibilidade ao atraso compartilhem a rede. Este algoritmo é uma variação do PQ, pois o buffer pode ser dividido em várias filas com diferentes níveis de preferência (classes). Entretanto, o algoritmo CBQ permite definir a quantidade de bytes de dados que deve ser escoada de cada fila do buffer para a fila da interface de saída do dispositivo de rede.

Utilizando esse algoritmo de fila de espera é possível prever uma dada largura de banda fixa num determinado ponto de rede congestionado para um conjunto de tráfego específico e deixar a banda remanescente para o tráfego restante. Isto é feito determinando uma quantidade específica de bytes da fila do buffer para cada classe de pacotes e depois servindo a fila da interface de saída utilizando o algoritmo de escalonamento **Round Robin**. Este algoritmo de escalonamento permite que cada fila do buffer seja acedida numa sequência pré-definida e uma dada quantidade de bytes seja extraída dessa fila. O resultado do acesso a essas filas do buffer numa sequência pré-definida cria diferentes níveis de prioridades entre estas. Logo a escolha da fila do buffer a ser acedida baseia-se na classe que esta fila pertence.

Weighted Fair Queuing (WFQ)

O algoritmo de fila de espera, traduzindo como justo e ponderado (Weighted Fair Queuing - WFQ) permite que cada fila do buffer tenha uma mesma quantidade de bytes extraída e direcionada para a fila da interface de saída do dispositivo de rede. Isto torna mais justo o escalonamento desses pacotes de dados e, conseqüentemente, equidade no fornecimento dos diferentes serviços. Além disso, o WFQ assegura que as filas dos buffers não sofram

por falta de largura de banda, pois a cada ciclo de execução desse algoritmo uma mesma quantidade de bytes é extraída de cada uma dessas filas. O termo ponderado empregado no nome deste algoritmo provém do fato desse utilizar “pesos” para determinar a ordem de enfileiramento dos pacotes na transmissão. Para a execução deste ordenamento é preciso que cada pacote seja diferenciado quanto seu nível de prioridade. Portanto, o WFQ analisa o campo Prioridade (Precedence) do cabeçalho do pacote IP a fim de atribuir um determinado peso a esse pacote. Esse campo contém uma prioridade que varia de 0 (normal) a 7 (pacote de controle de rede). Quanto maior o nível (valor) de prioridade menor é o valor do peso atribuído a esse pacote. Assim, os pacotes de dados com baixos pesos são escalonados primeiro, ou seja, garantem ao fluxo desses pacotes um tempo de resposta mais imediato, tornando este algoritmo mais adaptativo às mudanças nas condições de tráfego da rede. O algoritmo WFQ é um mecanismo eficiente na medida em que esse pode utilizar toda largura de banda disponível para transmitir o tráfego de baixa prioridade se não existir tráfego de mais alta prioridade.